BLM3590 - İstatistiksel Veri Analizi

Doç. Dr. Ali Can KARACA & Arş. Gör. Kübra ADALI

Yusuf Taha TÜTEN

21011027

taha.tuten@std.yildiz.edu.tr

Tel No: 532 574 55 66

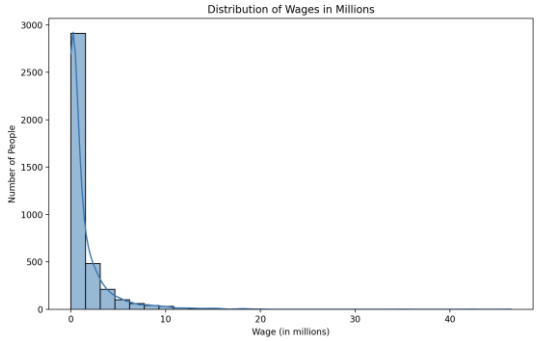
CONTENT

This report presents data on the annual salaries of footballers in Europe's six biggest leagues, collected from the Football Manager 22 game. While some individual values may be incorrect when compared to current values, the dataset generally provides accurate information on footballers' salaries.

Data Visualization

Number of People-Wage Histogram:

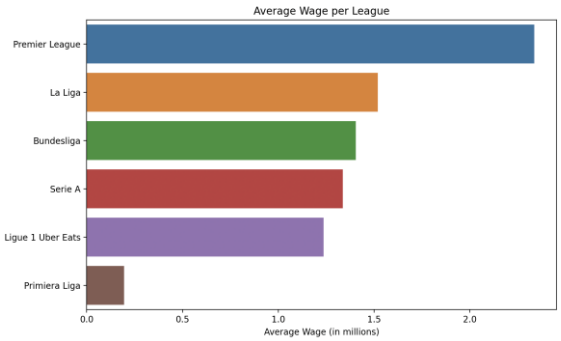
This histogram displays the number of individuals and their corresponding wages. It provides insight into the typical salary range for footballers.



Average Wage per League Bar Chart:

This bar chart displays the average wages of Europe's six largest football leagues: the Premier League (England), La Liga (Spain), Bundesliga (Germany), Serie A (Italy), Ligue 1 Uber Eats (France), and Primiera Liga (Portugal).

As shown here, the English league is ahead of others in terms of monetisation.



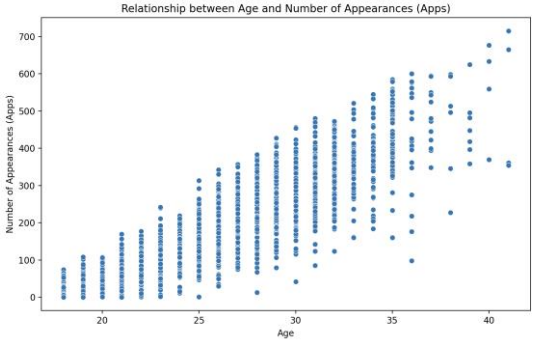
Age-Wage Scatter Plot:

The scatter plot depicts the correlation between age and wage, investigating whether wages increase with age. The report will further examine this relationship.

Age-Appearences Scatter Plot:

This scatter plot shows the number of games played by footballers throughout their careers and their ages, in order to determine if there is a correlation between these values.

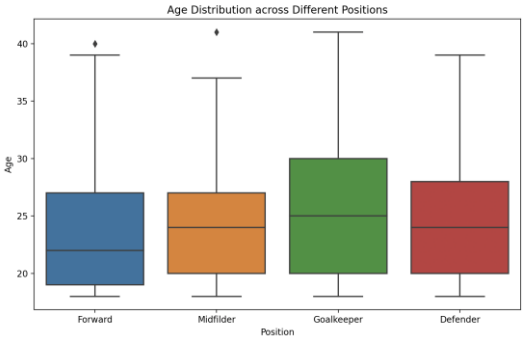
As seen here, there is a positive correlation between age and the number of appearances, as expected.



Age-Positions Box Plot:

This box plot displays the average and interquartile ages of football players grouped by their positions.

As seen here goalkeepers tend to play more than players in other positions, possibly due to injuries. This is because players in other positions are more active than goalkeepers.



Descriptive Statistics of Data

Wage:

Count: 3,907	Mean: $1,367,959	Standard Deviation: $2,589,857	Minimum: $1,400
25th Percentile: $75,500	Median: $399,000	75th Percentile: $1,560,000	Maximum: $46,427,000

Age:

Count: 3,907	Mean: 24.12 years	Standard Deviation: 4.94 years	Minimum: 18 years
25th Percentile: 20 years	Median: 24 years	75th Percentile: 28 years	Maximum: 41 years

Apps (Appearances):

Count: 3,907	Mean: 140.06	Standard Deviation: 131.69	Minimum: 0

25th Percentile: 15 apps	Median: 115	75th Percentile: 224 apps	Maximum: 715

Caps:

Count: 3,907	Mean: 8.93	Standard Deviation: 20.52	Minimum: 0
25th Percentile: 0 caps	Median: 0	75th Percentile: 6 caps	Maximum: 180

Hypothesis Test

Null Hyptohesis(H0): The players' average wage is 1,5 million. H0: $\mu = 1{,}500{,}000$

Alternative Hypothesis(H1): The players' average wage is lesser than 1,5 million H1: $\mu < 1{,}500{,}000$

The validity of our hypothesis will be determined using the z-test. To begin, a confidence level, such as 95%, must be selected. This means that our C value is 0.95 and our alpha value is 0.05.

To calculate the z-score under these circumstances, we have a sample mean (x^-) of 1,367,959, an expected mean (μ) of 1,500,000, and a standard deviation of sigma. The standard deviation (σ) was 2,589,857.

$$z - score = \frac{x^- - \mu}{\sigma}$$

The formula for z-score is used to calculate the value of –0.050983. This, in turn, gives us a p-value of 0.47967. As the p-value is greater than alpha, we can not reject the null hypothesis.

Correlation Analysis

Suppose we wish to calculate the correlation between a player's age and their wage to determine if older players earn more than younger ones. It is expected that a player will gain more popularity over time, which may result in a higher salary. To determine the validity of this claim, we need to calculate the Pearson correlation coefficient, t-statistics, and p-value to determine the level of confidence in our correlation.

Null Hyptohesis(H0): There is no correlation between age and wage.

Alternative Hypothesis(H1): There is correlation between age and wage.

$$r = \frac{\sum (x_i - x^-)(y_i - y^-)}{\sqrt{\sum (x_i - x^-)^2 \ \sum (y_i - y^-)^2}}$$

The Pearson correlation coefficient formula is as follows: X represents age variables and Y represents wage variables. The resulting correlation coefficient is r = 0.317.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

After calculating the correlation coefficient (r), it is necessary to calculate the p-value to determine the confidence of the correlation. Therefore, t statistics are required to calculate the p-value. With n=3,907 and r=0.317, we obtain t=18.53. After calculating T(18.53), we obtain a p-value of 2.603*10^-92. Therefore, we can conclude that this correlation coefficient is significant due to the extremely small p-value. So we reject can reject null hypothesis.

The correlation coefficient of 0.317 indicates a positive moderate correlation between age and wage variables. Therefore, a player's age has a moderate impact on their wage. Early-peaking superstars such as Haaland, Mbappe, and Bellingham are examples of the opposite.